The neutral theory in the genomic era

Justin C Fay*‡ and Chung-I Wu*†

A number of tests have been developed to detect positive selection at the molecular level. These tests are based on DNA polymorphism within and divergence between species. Applications of these tests have revealed a large collection of genes that have evolved under positive selection and some general insights into adaptive evolution. Recently, these tests have been applied on a genomic scale and have provided estimates of the frequency of adaptive substitutions and a critical test of the neutral theory.

Addresses
*Committee on Genetics, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA
†Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA
‡e-mail: jcfay@lbl.gov
Correspondence: Justin C Fay

Current Opinion in Genetics & Development 2001, **11**:642–646

0959-437X/01/$ – see front matter

Abbreviations

α	proportion of substitutions driven by positive selection
A	amino acid
β	fraction of amino acid polymorphism that is deleterious
ka	rate of amino acid substitution
ks	rate of synonymous substitution
MK	McDonald–Kreitman
N	effective population size
s	selection coefficient
S	synonymous

Introduction

The neutral theory of molecular evolution posits that the majority of DNA variation within and between species is neutral with respect to fitness and can be described by stochastic fluctuations in a finite population [1]. The formulation of the neutral theory provided a number of predictions that could be tested using patterns of DNA polymorphism and divergence [2]. There are two approaches. In the direct approach, amino acid changes are compared to synonymous changes. Because synonymous changes are neutral or nearly neutral with respect to fitness, any difference between patterns of amino acid and synonymous changes within and/or between species can be attributed to selection on amino acid changes. In the indirect approach, selection is detected on the basis of a skew in the frequency distribution of neutral variation linked to a site that has been under selection. The frequency spectrum can also be influenced by a population's demographic history. Positive selection can be distinguished from demographic effects because selection produces a local skew in the frequency spectrum whereas demography produces a genome-wide skew. Although these methods sometimes reveal evidence for positive selection in individual genes, the neutral theory can only be addressed with a genomic estimate of the frequency of positive selection. Recent studies of genomic patterns of amino acid and synonymous polymorphism and divergence have provided insights into the mode and, in some cases, tempo of adaptive evolution.

Divergence

A rate of amino acid substitution (*ka*) greater than that of synonymous substitution (*ks*) is the most robust test for positive selection [3]. A large collection of rapidly evolving genes have been identified by this criteria. The majority are involved in either sexual reproduction or host–pathogen interactions [4•]. This can be explained by the continual improvement in fitness required by sexual selection [5•], and the evolutionary 'arms race' between a host and its pathogen [6•]. The extent to which these genes are reflective of adaptive evolution, however, is not known as even a small amount of constraint on a protein can reduce $ka/ks < 1$.

The power of detecting positive selection is increased through analysis of different codons within a gene [7,8], of different classes of amino acid substitutions [5•,9], and along different phylogenetic lineages [10–12]. By assuming that *ka*/*ks* varies across sites, those codons that are likely to have been under positive selection can be identified [7,13•,14]. Whether approximate or maximum likelihood methods are used, it is clear that mutational and codon biases must be accounted for when estimating *ka*/*ks* [4•,15].

To address the frequency of positive selection in protein evolution, a genomic approach must be taken. For reproductive proteins there is pervasive evidence of positive selection. The rate of divergence of *Drosophila* male and female reproductive proteins is much greater than that of non-reproductive-associated genes [16,17]. Differences in the rate of protein divergence, however, may be caused by differences in selective constraint between the two classes of genes, unless *ka*/*ks* is shown to be >1. This was done in a recent study [18•], which found 11% (19/176) of male reproductive proteins have $ka/ks > 1$ between *D. melanogaster* and *D. simulans*. Previous studies have demonstrated that for some of these genes, the rapid rate of evolution can only be explained by positive selection [19,20]. Male reproductive proteins in primates have also been evolving rapidly; out of 18 male reproductive proteins, positive selection has increased the rate of conservative amino acid substitutions above the synonymous rate in the 11 most rapidly evolving genes [5•].

What fraction of all genes in the genome, not just male reproductive genes, show evidence of positive selection? Using all groups of homologous sequences available in public databases, 0.5% (17/3595) were shown [21] to have $ka > ks$ for more than half of the pairwise comparisons in

the group. More recently, 5.3% (280/5305) of chordate gene families and 3.6% (123/3385) of embryophyta gene families were found to contain at least one branch with a *ka*/*ks* value >1 [22]. More detailed analyses will be needed to place statistical significance on these results and estimate the rate of adaptive substitution.

Polymorphism

Patterns of polymorphism provide another means of detecting positive selection. The spread of an advantageous mutation through a population produces both a reduction in levels of linked neutral variation [23] and a skew in the frequency spectrum [24••]. Positive selection can be detected by a local reduction in levels of variation using the HKA test [25], which compares polymorphism and divergence between two or more regions, or on a genomic level by a positive correlation between levels of variation and rates of recombination. This correlation has been found in a number of species and implies that selection is frequent enough to eliminate most variation from regions of low recombination within the genome (see the review by Aquadro *et al*., this issue [pp 627–634]); however, the elimination of linked neutral variation as a result of background selection against deleterious mutations provides an alternative explanation of the data [26].

The frequency spectrum can be used to distinguish background selection from hitchhiking [27]. Hitchhiking sweeps variation to low and high frequencies [24••], whereas background selection is not expected to produce a skew in the frequency spectrum, at least for small samples sizes [28]. Compared to intermediate frequency variation, both an excess of low frequency variation, as measured by the D statistic, and an excess of high frequency variation, as measured by the H statistic, can be used to detect hitchhiking [24••,29]. Using the H statistic, one out of three regions of low recombination in *D. melanogaster* was found to have a significant excess of high-frequency variants [24••]. Additional evidence for hitchhiking comes from the correlation found between rates of recombination and the D statistic [30•], as is expected under a model of recurrent hitchhiking [31]. Because the combined action of both positive and negative selection cannot easily be disentangled [32•], it would be difficult to obtain an estimate of the frequency of adaptive substitutions on the basis of a reduction in levels of variation [33].

With large-scale surveys of polymorphism, positive selection can be detected by scanning regions of high recombination for a local reduction in levels of neutral variation accompanied by a skew in the frequency spectrum. This method has been used in the MHC region of humans [34] and in the *D. melanogaster* genome using microsatellite variation [35,36]. Similarly, a local reduction in levels of variation surrounding the newly evolved *Sdic* gene in *D. melanogaster* was used as evidence for positive selection at this locus [37]. Linkage disequilibrium mapping uses levels of phenotypic and genotypic associations to simultaneously detect selection and map the mutation responsible for the phenotype. This has been used to map the warfarin resistance gene in rats [38••], and the desaturase gene responsible for a cuticular hydrocarbon pheromone polymorphism in *D. melanogaster* [39•]. Linkage disequilibrium has been examined on a genomic scale in humans to identify regions that are likely to have been influenced by selection [40].

Table 1

Estimates of the fraction of low frequency amino acid polymorphism that is deleterious (β) and the fraction of amino acid substitutions that were driven by positive selection (α).

Data	A/S			β	α	Refs.
	Rare	Common	Fixed			
H. sapiens	1.11	0.57	0.88	0.48	0.35	[42•]
D. melanogaster	0.63	0.29	0.63	0.54	0.54	(a)
D. simulans		0.06	0.20		0.70	[57]
D. simulans complex		0.07	0.21		0.68	[58]

(a) JC Fay *et al*., unpublished data.

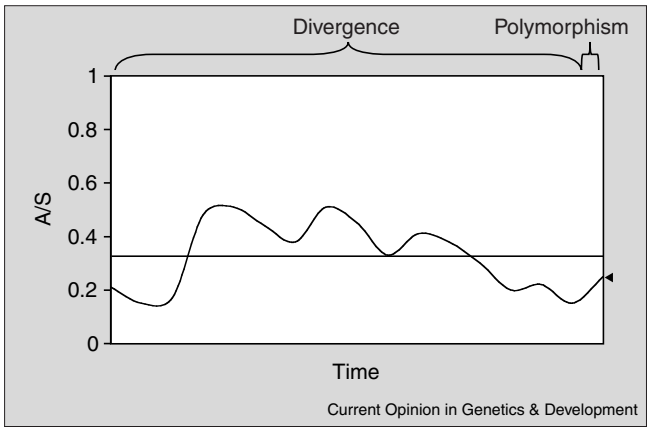
Positive selection can also be detected by a skew in the frequency spectrum of amino acid compared to synonymous variation. The largest polymorphism survey with outgroup sequence [41], which is necessary to distinguish low and high frequency mutations, shows a larger A/S ratio at high (7/7) compared to intermediate frequencies (28/38), as expected if some amino acid mutations were advantageous [42•]. With more data, the strength and frequency of adaptive mutations could be estimated.

Polymorphism and divergence

The McDonald–Kreitman (MK) test has the advantage of being able to detect positive selection in the presence of a strong selective constraint [43]. The test compares the ratio of the number of amino acid (A) to synonymous (S) polymorphic sites to the A/S ratio of fixed differences between species. If a large fraction of amino acid substitutions are driven by positive selection, the A/S ratio of divergence should be inflated above that of polymorphism. Although the power of the MK test is clearly much greater than *ka*/*ks* [44], it is based on several assumptions that are often not justified. By comparing polymorphism and divergence from a large number of genes these assumptions can be either removed or tested, as discussed below.

The proportion of amino acid substitutions driven by positive selection can be estimated from the difference between the A/S ratio of common polymorphism and divergence (the use of common polymorphism is discussed below). For all species with data from a large number of genes, the estimate is surprisingly high (Table 1); however, these estimates assume synonymous mutations are neutral and the constraint on a gene remains constant over time. Although there is evidence for selection on synonymous

Figure 1



Hypothetical ratio of neutral amino acid (A) to synonymous (S) mutations that become fixed along a lineage. The A/S ratio of divergence is a measure of the average amount of constraint on a gene and is the cumulative number of amino acids divided by synonymous substitutions over time, as shown by the horizontal line. The A/S ratio of polymorphism is a measure of the selective constraint on a gene in the recent past, as shown by the arrowhead.

sites in *Drosophila* [45], codon bias is not correlated with the rate of synonymous substitution [46••,47]. This indicates that a change in the strength of selection on synonymous sites would be too weak to produce a change in the A/S ratio. A change in selective constraint on the protein sequence is a more serious consideration.

The assumption that the selective constraint on a gene is constant, as measured by A/S, is rarely justified. The constraint on a gene is determined by the fraction of amino acid altering mutations for which $2Ns < -1$, where N is the effective population size and s is the selection coefficient [3]. Fluctuations in either population size or the strength of selection are expected to produce fluctuations in the A/S ratio of mutations that become fixed along a lineage (Figure 1) [48,49]. By chance, the A/S ratio of extant polymorphism may be lower than that of divergence, which is a measure of the average A/S ratio between two species. Although fluctuations in the intensity of selection may either increase or decrease the A/S ratio of a single gene over time, only a change in effective population size is expected to produce a genome-wide effect on the A/S ratio. Regardless of the cause, the magnitude of the fluctuation in the A/S ratio depends on the distribution of Ns values associated with amino acid mutations. If a large proportion of mutations have Ns values that lie close to –1, then the A/S ratio is expected to be quite sensitive to even slight changes in N or s.

The nearly neutral theory of molecular evolution places an emphasis on the role of slightly deleterious amino acid mutations in explaining patterns of DNA variation [50]. The original support for the theory came from the large excess of allozyme variation found at low frequencies within human and *Drosophila* populations [51]. Demographic explanations have now been ruled out by showing there is an excess of low frequency amino acid compared to synonymous polymorphism ([42•]; JC Fay *et al*., unpublished data). The difference between the A/S ratio of rare compared to common polymorphism was used to obtain estimates in both *Homo sapiens* (181 genes) and *D. melanogaster* (31 genes) of the fraction of amino acid polymorphism which is slightly deleterious (Table 1). Slightly deleterious amino acid mutations may lie approximately between $-100 < 2Ns < -1$, but are simply defined as those mutations with significant effects on levels of polymorphism (>1%) and rates of divergence. By fitting the data to a population genetic model, it was estimated that 20–40% of amino acid mutations are slightly deleterious in humans and the average individual is expected to carry over 500 of these mutations in their genome [42•].

The large fraction of amino acid mutations that have been estimated to be slightly deleterious have two important implications for applications of the MK test. The first is that the A/S ratio of polymorphism is inflated by slightly deleterious amino acid mutations segregating at low frequencies in a population. This effect can be removed, at least in a crude manner, by comparison of the A/S ratio of common polymorphism with divergence [42•,45,52]. The second implication is that the A/S ratio may be particularly susceptible to changes in effective population size. Given a reduction in effective population size, the A/S ratio of subsequent substitutions may approach that of low-frequency variation, which is quite high (Table 1). Therefore, a high A/S ratio of divergence may be accounted for by a period of reduced effective population size.

A change in population size can be distinguished from positive selection by comparison of the A/S ratio of rare and common polymorphism to that of fixed differences (JC Fay *et al*., unpublished data). A change in effective population size should affect all genes, and the degree to which each gene is affected should depend on the fraction of amino acid mutations in that gene which are slightly deleterious. Thus, a change in population size predicts a correlation between the A/S ratio of divergence and that of low-frequency polymorphism. On the other hand, positive selection should not affect all genes to the same extent. This approach was taken in the analysis of polymorphism in *D. melanogaster* and divergence to *D. simulans* (JC Fay *et al*., unpublished data). The difference between 11 rapidly evolving genes, which were found to account for the entire difference between the A/S ratio of polymorphism and divergence, and 34 slowly evolving genes was found to be compatible with positive selection but not a change in effective population size. The slowly evolving genes were found to have an excess of low frequency amino acid variation but only a slight excess of amino acid divergence, whereas the rapidly evolving genes were found to have a large excess of both low-frequency polymorphism and divergence compared to common polymorphism (JC Fay *et al*., unpublished data).

Additional theoretical and empirical research will be needed to refine estimates of the fraction of amino acid mutations that are both deleterious, neutral and adaptive. The human polymorphism and divergence data contain different genes and the *Drosophila* data are limited by inconsistent sampling. An unbiased sample of 50 or more genes from a species whose history is unrelated to human demography is needed. The comparison of the A/S ratio of low and common polymorphism to that of fixed differences allows for only a rough estimate of the impact of slightly deleterious and adaptive mutations on protein polymorphism and divergence. Optimally, a maximum likelihood method, which incorporates mutational and codon usage biases, should be used in the analysis of polymorphism and divergence data [53]. Although some theoretical work has been done [48,49,54,55•,56], empirically the effect of a change in population size on the rate of molecular evolution is not well understood but is most pertinent to our understanding of molecular evolution.

Conclusions

The analysis of DNA polymorphism and divergence on a genomic scale will provide critical tests of the neutral theory of molecular evolution. Application of the MK and other A/S ratio based tests have provided substantial evidence that positive selection has inflated the rate of protein divergence above that expected on the basis of polymorphism. Future studies will need to estimate the relative contributions of changes in population size and positive selection to this pattern. The role of positive selection in the divergence of duplicate genes and in the evolution of gene regulatory elements has yet to be addressed.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kimura M: *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press; 1983.

2. Kreitman M, Akashi H: **Molecular evidence for natural selection.** *Annu Rev Ecol Syst* 1995, **26**:403-422.

3. Kimura M: **Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution.** *Nature* 1977, **267**:275-276.

4. • Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation.** *Trends Ecol Evol* 2000, **15**:496-503.
This review provides a concise and clear description of maximum likelihood methods used to detect positive selection. Examples and a summary of genes for which there is evidence of positive selection is also presented.

5. • Wyckoff GJ, Wang W, Wu CI: **Rapid evolution of male reproductive genes in the descent of man.** *Nature* 2000, **403**:304-309.
Although many examples of genes under positive selection are known, in this work an entire class of genes involved in male reproduction show evidence of positive selection. To amplify the signal of positive selection, conservative and radical amino acid substitutions are distinguished and simulations are used to show there is a significant excess of conservative changes found in primate male reproductive genes compared to a reference set of genes.

6. • Bergelson J, Kreitman M, Stahl EA, Tian D: **Evolutionary dynamics of plant R-genes.** *Science* 2000, **292**:2281-2285.
Although much progress has been made in understanding the molecular basis of pathogen virility and plant resistance, the evolutionary forces driving this co-evolution are not well understood. By comparing rates of evolution between paralogous and orthologous *Arabidopsis thaliana* R-genes, evidence for positive selection is found.

7. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**:929-936.

8. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431-449.

9. Yang Z, Nielsen R, Hasegawa M: **Models of amino acid substitution and applications to mitochondrial protein evolution.** *Mol Biol Evol* 1998, **15**:1600-1611.

10. Yang Z, Nielsen R: **Synonymous and nonsynonymous rate variation in nuclear genes of mammals.** *J Mol Evol* 1998, **46**:409-418.

11. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Mol Biol Evol* 1998, **15**:568-573.

12. Zhang J, Rosenberg H, Nei M: **Positive Darwinian selection after gene duplication in primate ribonuclease genes.** *Proc Natl Acad Sci USA* 1998, **95**:3708-3713.

13. • Yang Z, Swanson WJ, Vacquier VD: **Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites.** *Mol Biol Evol* 2000, **17**:1446-1455.
Significant variation in selective pressure is shown along abalone lineages and both constrained and adaptive sites within the lysin protein are identified. An excellent example of application of recently developed maximum likelihood methods.

14. Suzuki Y, Gojobori T: **A method for detecting positive selection at single amino acid sites.** *Mol Biol Evol* 1999, **16**:1315-1328.

15. Yang Z, Nielsen R: **Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models.** *Mol Biol Evol* 2000, **17**:32-43.

16. Civetta A, Singh RS: **High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species.** *J Mol Evol* 1995, **41**:1085-1095.

17. Coulthart MB, Singh RS: **High level of divergence of male reproductive tract proteins between *Drosophila melanogaster* and its sibling species, *D. simulans*.** *Mol Biol Evol* 1988, **5**:182-191.

18. • Swanson W, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF: **Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*.** *Proc Natl Acad Sci USA* 2001, **98**:7375-7379.
Using differential hybridization, male accessory gland proteins are identified and sequenced from *D. simulans.* These genes have diverged more than twice as fast as non-reproductive proteins and show evidence of positive selection.

19. Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG: **Molecular population genetics of male accessory gland proteins in *Drosophila*.** *Genetics* 2000, **156**:1879-1888.

20. Tsaur SC, Wu CI: **Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*.** *Mol Biol Evol* 1997, **14**:544-549.

21. Endo T, Ikeo K, Gojobori T: **Large-scale search for genes on which positive selection may operate.** *Mol Biol Evol* 1996, **13**:685-690.

22. Liberles DA, Schreiber DR, Govindarajan S, Chamberlin SG, Benner SA: **The adaptive evolution database (TAED).** *Genome Biol* 2001, **2**:0028.

23. Maynard Smith J, Haigh J: **The hitch-hiking effect of a favorable gene.** *Genet Res* 1974, **23**:23-35.

24. •• Fay JC, Wu CI: **Hitchhiking under positive Darwinian selection.** *Genetics* 2000, **155**:1405-1413.
Here, we show hitchhiking produces a skew in the frequency spectrum with the unique feature being an excess of high-frequency mutations. A statistical test is developed on the basis of this pattern and is shown to be useful in both regions of low and high recombination.

25. Hudson RR, Kreitman M, Aguade M: **A test of neutral molecular evolution based on nucleotide data.** *Genetics* 1987, **116**:153-159.

26. Charlesworth B, Morgan MT, Charlesworth D: **The effect of deleterious mutations on neutral molecular variation.** *Genetics* 1993, **134**:1289-1303.

27. Charlesworth D, Charlesworth B, Morgan MT: **The pattern of neutral molecular variation under the background selection model.** *Genetics* 1995, **141**:1619-1632.

28. Fu YX: **Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.** *Genetics* 1997, **147**:915-925.

29. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**:585-595.

30. • Andolfatto P, Przeworski M: **Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*.** *Genetics* 2001, **158**:657-665.

Although previous studies found no assocation between recombination rate and a skew in the frequency spectrum [27], a significant correlation is found when examining African populations of *D. melanogaster*. This report illustrates why non-African populations may not always be suitable for population genetic analyses because of their complex demographic history.

31. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W: **The hitchhiking effect on the site frequency spectrum of DNA polymorphisms.** *Genetics* 1995, **140**:783-796.

32. • Kim Y, Stephan W: **Joint effects of genetic hitchhiking and background selection on neutral variation.** *Genetics* 2000, **155**:1415-1427.

The first analytical and simulation results on the combined effects of background selection and hitchhiking on patterns and levels of DNA variation. The result is that in regions of very low recombination, hitchhiking has the predominant effect on levels of variation whereas the opposite is true in regions of higher recombination.

33. Wiehe TH, Stephan W: **Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*.** *Mol Biol Evol* 1993, **10**:842-854.

34. Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T: **SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity.** *Genome Res* 2000, **10**:1579-1586.

35. Irvin SD, Wetterstrand KA, Hutter CM, Aquadro CF: **Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*. Evidence for founder effects in new world populations.** *Genetics* 1998, **150**:777-790.

36. Schlotterer C, Vogl C, Tautz D: **Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations.** *Genetics* 1997, **146**:309-320.

37. Nurminsky D, Aguiar DD, Bustamante CD, Hartl DL: **Chromosomal effects of rapid gene evolution in *Drosophila melanogaster*.** *Science* 2001, **291**:128-130.

38. •• Kohn MH, Pelz HJ, Wayne RK: **Natural selection mapping of the warfarin-resistance gene.** *Proc Natl Acad Sci USA* 2000, **97**:7911-7915.

Because selection creates linkage disequilibrium, linkage disequilibrium can be used to pinpoint genes under strong selection. In the case of warfarin-resistance, linkage disequilibrium was used for both mapping and understanding the evolutionary history of diseases resistance. This study provides the first look into the origin and population dynamics of warfarin-resistance at the molecular level.

39. • Takahashi A, Tsaur SC, Coyne JA, Wu CI: **The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2001, **98**:3920-3925.

A combination of recombination and association mapping techniques were used to map a naturally occurring cuticular hydrocarbon polymorphism to a 16bp deletion 5¢ of a desaturase gene. The deletion distinguishes African and Caribbean from other world populations and shows evidence of recent positive selection by an excess of high frequency variants in its vicinity. The polymorphism is of particular interest as it may be involved in sexual selection and the premating isolation found between African and non-African populations.

40. Huttley GA, Smith MW, Carrington M, O'Brien SJ: **A scan for linkage disequilibrium across the human genome.** *Genetics* 1999, **152**:1711-1722.

41. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N *et al.*: **Characterization of single-nucleotide polymorphisms in coding regions of human genes.** *Nat Genet* 1999, **22**:231-238.

42. • Fay JC, Wyckoff GJ, Wu CI: **Darwinian selection on the human genome.** *Genetics* 2001, **158**:1217-1226.

We confirm Ohta's 1975 [51] finding that a large fraction of amino acid mutations are slightly deleterious is confirmed. The frequency and average number of slightly deleterious mutations per genome is estimated. The rate of amino acid divergence between human and old world monkey is found to be higher than expected on the basis of polymorphism.

43. McDonald JH, Kreitman M: **Adaptive protein evolution at the *Adh* locus in *Drosophila*.** *Nature* 1991, **351**:652-654.

44. Akashi H: **Within- and between-species DNA sequence variation and the 'footprint' of natural selection.** *Gene* 1999, **238**:39-51.

45. Akashi H: **Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift.** *Gene* 1997, **205**:269-278.

46. •• McVean GA, Vieira J: **Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*.** *Genetics* 2001, **157**:245-257.

Both population genetic and phylogenetic models are incorporated into a maximum likelihood method by which the selection, mutation and demographic parameters are estimated in a comprehensive analysis of codon bias. It is shown that codon bias can have a substantial impact on estimates of divergence.

47. Marais G, Mouchiroud D, Duret L: **Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes.** *Proc Natl Acad Sci USA* 2001, **98**:5688-5692.

48. Araki H, Tachida H: **Bottleneck effect on evolutionary rate in the nearly neutral mutation model.** *Genetics* 1997, **147**:907-914.

49. Tachida H: **DNA evolution under weak selection.** *Gene* 2000, **261**:3-9.

50. Ohta T: **Slightly deleterious mutant substitutions in evolution.** *Nature* 1973, **246**:96-98.

51. Ohta T: **Statistical analyses of *Drosophila* and human protein polymorphism.** *Proc Nat Acad Sci USA* 1975, **72**.

52. Templeton AR: **A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus.** *Genetics* 1995, **140**:403-409.

53. Hasegawa M, Cao Y, Yang Z: **Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species.** *Mol Biol Evol* 1998, **15**:1499-1505.

54. Gillespie JH: **The neutral theory in an infinite population.** *Gene* 2000, **261**:11-18.

55. • Gillespie JH: **The role of population size in molecular evolution.** *Theor Popul Biol* 1999, **55**:145-156.

The effect of changes in population size on a range of evolutionary models is found using computer simulations. The distribution of selection coefficients and their interactions are shown to be critical to understanding the role of effective population size in molecular evolution.

56. Cherry JL: **Should we expect substitution rate to depend on population size?** *Genetics* 1998, **150**:911-919.

57. Begun DJ: **The frequency distribution of nucleotide variation in *Drosophila simulans*.** *Mol Biol Evol* 2001, **18**:1343-1352.

58. Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J: **The population genetics of the origin and divergence of the *Drosophila* simulans complex species.** *Genetics* 2000, **156**:1913-1931.